PREDICTION AND COMPARISON OF HIV-1 PROTEASE INHIBITOR BINDING ENERGIES BY VARIOUS MOLECULAR DOCKING METHODS

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology
in
Biochemical Engineering and Biotechnology

By

Koteswara Reddy Gujjula

(20600003)

Department of Chemical Engineering

National Institute of Technology

Rourkela, Orissa

May-2008

PREDICTION AND COMPARISON OF HIV-1 PROTEASE INHIBITOR BINDING ENERGIES BY VARIOUS MOLECULAR DOCKING METHODS

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology
in
Biochemical Engineering and Biotechnology

Submitted by

Koteswara Reddy Gujjula

Under the guidance of

Prof. Gyana R. Satpathy

Department of Biotechnology & Medical Engineering

National Institute of Technology

Rourkela, Orissa

May-2008

National Institute of Technology, Rourkela
Orissa-769008, India

CERTIFICATE

This is to certify that the thesis entitled, " Prediction and comparison of HIV-1 Protease inhibitor binding energies by various molecular docking methods" submitted by *Mr. Koteswara Reddy Gujjula* in partial fulfillment of the requirements for the award of Master of Technology in Chemical Engineering with specialization in " Biochemical Engineering and Biotechnology" at the National Institute of Technology, Rourkela (Deemed University) is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma.

Date:

Prof. Gyan R.Satpathy
Dept. of Biotechnology & Medical Engg.,
National Institute of Technology,
Rourkela – 769008.

ACKNOWLEDGEMENT

I wish to express my sincere thanks to my guide, Prof.Gyan R.Satpathy for his able guidance and advices during my project work. Thank you for your patience and understanding. I must also acknowledge the HOD, Prof. K.C. Biswal for given this excellent opportunity to complete it successfully.

I am also very grateful to Prof. B.N.Misra (ACS-Bioinformatics, Biotech Park,Lucknow) and his group for training in structural bioinformatics techniques. And I would like to say thanks to my friends Mr.S.C.Patnaik, Mr.Shadrack Jabes, and Mr.E.H.K.Reddy those who are directly or indirectly supported me in carrying out this thesis work successfully.

Koteswara Reddy Gujjula
M.Tech, Biochemical Engineering and Biotechnology
Roll no: 2060003
N.I.T. Rourkela-769008.

ABSTRACT

The human immunodeficiency virus type 1 *Asp*artic protease (HIV-1 PR) is an important enzyme due to its vital role in viral maturation. Inactivation of the enzyme causes the production of immature viral particles. The accurate prediction of enzyme-substrate interaction energies is one of the major challenges in computational biology. Docking experiments were undertaken using the programs AutoDock.4 and online programs pardock for twenty-five HIV-1 protease-inhibitor complexes determined by x-ray crystallography. From the molecular docking study, we were able to select a best solution based on lowest binding energy and lowest RMSD values of receptor-ligand complex in each docking program.

Correlations observed for experimental and predicted binding energy values for receptor-ligand complex. A highest correlation coefficient of 0.801 was observed between the experimental and predicted binding energy for pardock program and 0.484 by autodock 4.0. Patch dock followed by firedock methods also used to predict the global energy of each enzyme-inhibitor complex and complementarily score. Our result indicates that the binding energies predicted by pardock program are highly correlated with experimental binding energies. The consensus ranking of enzyme-inhibitor complexes in various molecular docking methods improve the binding energy predictions. Consensus ranking has become an important method in various molecular docking methods to identify new inhibitors in computer-assisted discovery of new pharmaceutics.

Keywords: HIV-1Protease, Computer Assisted Drug Design, Molecular docking, Autodock4.0, Pardock.

Abbreviations

ACE	Atomic Contact Energy
AIDS	Acquired Immune Deficiency Syndrome
BFE	Binding Free Energy
CA	Capsid protein
CDC	Center for Diseases Control
CD4	Receptor found on surface of certain immune cells, Cluster of Differentiation antigen 4
COM	Center Of Mass
DNA	Deoxyribo Nucleic Acid
DPF	Docking Parameter File
FDA	U.S Food and Drug Administration
gp41, gp120	*Gly*coprotein's 41,120
HAART	Highly Active Anti-Retroviral Therapy
GPF	Grid Parameter File
HIV-1	Human Immunodeficiency Virus 1
HTS	High *Thr*oughput Screening
LGA	Lamarckian Genetic Algorithm
NAMD	Nano Molecular Dynamics
NNIs	Non-Nucleoside Inhibitors
PDB	Protein Data Bank
PR	Protease
PS	Pico Seconds
RCSB	Research Collaboratory for Structural Bioinformatics
RMSD	Root-Mean-Square-Distance
RNA	Ribo Nucleic Acid
RT	Reverse Transcriptase
SA	Simulated Annealing
UNAIDS	Joint United Nations program on HIV and AIDS
VDW	Van Der Walls forces

List of figures

List of tables

List of plots

Contents

Chapter 1

INTRODUCTION

1.1 Human immunodeficiency virus (HIV)

1.2 HIV-1 Protease

1.3 Protein Data Bank (PDB) files

1.4 Protein-ligand interactions

1.5 Binding energy

1.1 Human immunodeficiency virus (HIV)

It has now been two decades since acquired immunodeficiency syndrome (AIDS) was first reported by the US Center for Diseases Control (CDC). A few years later, it was found that a retrovirus called human immunodeficiency virus (HIV) is the causative agent in AIDS [1]. Since the outbreak of the AIDS epidemic, tremendous efforts directed towards development of antiretroviral therapies that target HIV type 1 in particular (HIV-1). Critical step in the life cycle of HIV is the proteolytic cleavage of the polypeptide precursors into mature enzymes and structural proteins catalyzed by HIV PR. It has been shown that budded immature viral particles that contain catalytically inactive protease cannot undergo maturation to an infective Form.2 The necessity of this enzyme in the virus life cycle makes it a promising target for therapy of the HIV infection[3].

1.2 HIV-1 Protease: HIV protease: a logical target for AIDS therapy

Highly conserved catalytic sequence: *Asp*-*Thr*-*Gly* (25-26-27) in each monomer [8]. That constitutes in part the substrate-binding site and plays an important role in substrate binding, and one of the two essential *Asp*artyl residues, *Asp*-25 and *Asp*-25 that lie on the bottom of the cavity. The substrate binds in its extended conformation, in which its interactions with the different amino acid side chains determine the specificity of the enzyme [7]. The two S1 subsites are very hydrophobic; the S2 subsites are mostly hydrophobic except *Asp*-29, *Asp*-29_, *Asp*-30 and *Asp*-30. The S3 subsites are adjacent to S1 subsites and are mostly hydrophobic. Antiretroviral drugs are developed for inactivation of hiv-1 protease. The accurate prediction of enzyme-substrate interaction energies is one of the major challenges in computational biology. For this, various docking methods have been developed with different force field parameters and scoring functions.

1.3 PDB files

HIV-1 protease-inhibitor complexes taken from Protein Data Bank in the form of PDB format. Since our interests are concentrated on small non-covalently bound ligands, those complexes containing covalently bound ligands, complex ligands and macromolecular ligands are stripped out of the data set.UCSF Chimera is a highly extensible, interactive molecular visualization and analysis system. Chimera can read molecular structures and associated data in a large number of formats, display the structures in a variety of

representations, and generate high-quality images and animations suitable for publication and presentation. Separated protein file and ligand file by gripping from protein-ligand complex.

Binding free energy was determined by differences in enthalpy and entropy of the free and bound state. The binding constant Ki assuming thermodynamic equilibrium conditions for the protein-ligand complex formation quantifies usually binding affinity. Given recent improvements in search algorithms and energy functions, computational docking methods have become a valuable tool to probe the interaction between an enzyme and its inhibitors. These methods can contribute significantly to the understanding of structural and energetic basis of enzyme-substrate interactions. Protein-ligand docking methods aim to predict the binding energy of the protein-ligand complex given the atomic coordinates. A set of HIV-1 protease-inhibitor binding energies were determined by various docking methods.

1.4 Protein-ligand interactions

Hydrogen bonds (e.g. C=O ... H-N): Hydrogen bonding is no doubt one of the key features for a specific binding process. Such interaction may happen when two atoms get close enough and form a donor-acceptor pair.

Hydrophobic interaction: --CH3------CH3-----

Hydrophobic Patch on the Inside of the Flap**.** Ile-47, Ile-54 and Val-56 are involved in hydrophobic interactions, With Ile-47 sitting just above Val-32 in the hydrophobic. Desolvation effect: Since both the protein and its ligand are solvated before complexation, the protein-ligand binding process was accompanied by desolvation, which undergoes changes in entropy as well as in enthalpy. This kind of effect is very difficult to characterize accurately. Both the long-range "Hydrophobic effect" and specific hydrogen bondings of water could be important in elucidation. These features make it nameable to simple additive pair wise interactions. So far, this effect was measured by calculating the buried hydrophobic surface areas during the binding process.

Van der Waals (VDW) interaction: This kind of interaction is a balance between attractive dispersion force and short-range repulsion. Although it was well accepted that van der Waals interactions play a fundamental role in the binding of the protein and its ligand, arguments exist in how to represent it in calculating the binding affinity. Some researchers assume that protein-ligand, protein-solvent, and ligand-solvent interfaces are

well packed and hence neglect any change in the VDW interactions upon binding. Some others assume that VDW interactions are better in a complex and therefore explicitly include them. By analyzing the training set, we believe that they all tell only part of the story. In general, one will find a closely packed interaction interface in a protein, ligand complex where many atom pairs are in a distance much shorter than the sum of their VDW radii, i.e. they form VDW bumps. Not all of these bumps come from hydrogen bonding or other strong interactions. Some of them are just the result of the tight binding between other parts of the protein and its ligand. It is not reasonable to assume that such a situation can also be found on the protein-solvent or ligand-solvent interface where the water molecules are mobile. Thus, our conclusion is that the VDW attraction between the protein and its ligand neglected due to the competitive interaction with water in the unbound state while, the VDW repulsion cannot. In the algorithms, the term for VDW interaction is simply a pair wise counting of VDW bumps between the protein and the ligand.

Metal-ligand bonding:

A variety of proteins have metal cations in their active sites, such as Mg2+, Ca2+, Mn2+, and Zn2+. In such cases, coordinate bonding between the metal and the ligand can often be important.

Deformation effect: Deformation refers to the conformational changes during the binding process. On one hand, it causes adverse entropic changes due to the freezing of internal rotations of both the protein and its ligand; on the other hand, it causes adverse enthalpy change due to the strain energy exerted during binding. Based on the principles of statistical thermodynamics, the entropic changes usually estimated by using a constant value per rotatable bond that is frozen. However, the enthalpy change is more difficult to elucidate.

1.5 Binding energy

Gibbs free energy of binding is ΔG then related to the binding constant by $\Delta G_0 = -RT \ln K_i$ At best, ΔG is determined by statistical thermodynamics resulting in a master equation that considers all contributing effects. More recently, empirical schemes have met with significant interest. The basic assumption underlined in such approaches is that the overall

binding free energy can be decomposed into components. This can be written out conceptually by the following equation.

$$\Delta G_{Binding} = \Delta G_{Motion} + \Delta G_{Interaction} + \Delta G_{Solvent} + \Delta G_{Cofiguration}$$

The parameters in the equation are often determined from binding data in a statistical manner. This kind of approach is also referred to as “Master Equation”. Computationally top-scoring solutions taken a selection from among the top scoring solution (e.g. the one closest to the experimental data among the ten top solutions) or simply the solution closest to the experimental data[40] (easiest and least rigorous approach). The accurate prediction of enzyme-substrate interaction energies is one of the major challenges in computational biology. Generally various docking methods followed by various energy scoring functions.

Chapter 2

LITERATURE REVIEW

2.1 Targeting HIV-1 protease

2.2 Motivation

2.3 The life cycle of HIV

2.4 Structure of HIV protease

2.5 Active site

2.6 Secondary structure

2.7 Mechanism of the HIV protease

2.7.1 Catalytic mechanism for HIV

2.1 Targeting HIV-1 protease

The human immunodeficiency virus type 1 aspartic protease (HIV-1 PR) is an important enzyme due to its vital role in viral maturation. Inactivation of the enzyme causes the production of immature viral particles. The accurate prediction of enzyme-substrate interaction energies is one of the major challenges in computational biology [1].

The enzyme therefore is an attractive target in anti-AIDS drug design, and the effect of binding various inhibitors on the protease, structure is currently the focus of intensive research. To obtain information about the position and energy of binding between an inhibitor and the corresponding protein, several automated docking programs have been developed (Ram Samudrala *et al.,* 2003).

2.2 Motivation

According to (Ram Samudrala *et al.,* 2003) improved the prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations than prediction of binding energies without any molecular dynamics simulations [1]. However, in computational biology using molecular dynamics simulation supercomputing facilities are needed. This computational method takes time-consuming process. We have focused on this method, and applied various molecular docking methods to predict the binding energies between hiv-1 protease-inhibitors.

One of these targets is HIV-1 protease (HIV PR), an essential enzyme needed in the proper assembly and maturation of infectious virions. Understanding the chemical mechanism of this enzyme has been a basic requirement in the development of efficient inhibitors.

2.3 The life cycle of HIV

HIV belongs to the class of viruses called retroviruses, which carry genetic information in the form of RNA. HIV infects T cells that carry the CD4 antigen on their surface. The infection of the virus requires fusion of the viral and cellular membranes, a process that mediated by the viral envelope *Gly*coprotein (gp120, gp41) and receptors (CD4 and co receptors, such as CCR5 or CXCR4) on the target cell. As the virus enters a cell, its RNA is reverse-transcribed to DNA by a virally encoded enzyme, the reverse transcriptase (RT). The viral DNA enters the cell nucleus, where it integrated into the genetic material of the cell by a second virally encoded enzyme, the integrase. Activation of the host cell results in

the transcription of the viral DNA into messenger RNA, which then translated into viral proteins. HIV protease, the third virally encoded enzyme, is required in this step to cleave a viral polyprotein precursor into individual mature proteins [2]. The viral RNA and viral proteins assemble at the cell surface into new virions, which then bud from the cell and released to infect another cell. The extensive cell damage from the destruction of the host's genetic system to the budding and release of virions leads to the death of the infected cells [2].

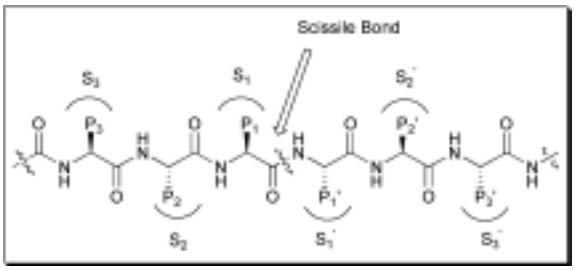
2.4 Structure of HIV protease

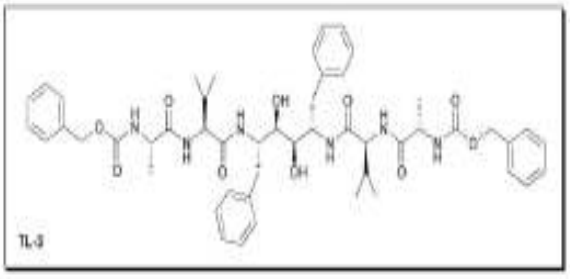
Navia *et al.* from Merck laboratories were the first group to obtain a crystal structure of HIV PR [4]. Kent and coworker reported a more accurate structure subsequently [5]. HIV PR is a 99 amino acid *Asp*artyl protease that functions as a homodimer with only one active site, which is *C***2**-symmetric in the free form. More than 140 structures of the HIV-1 PR, its mutants and enzymes complexed with various inhibitors reported so far. A database dedicated to providing structural information about HIV PR created at the National Cancer Institute (http://www-fbsc.ncifcrf.gov/HIVdb). The enzyme homodimer complexed with TL-3[6] shown in Fig.1 (PDB ID: 3TLH). Each monomer contains an extended β-sheet region (a *Gly*cine-rich loop) known as the flap, that constitutes in part the substrate-binding site and plays an important role in substrate binding, and one of the two essential *Asp*artyl residues, *Asp*-25 and *Asp*-25_ which lie on the bottom of the cavity. The substrate binds in its extended conformation, in which its interactions with the different amino acid side chains determine specificity of the enzyme [7].

Fig.1 Structure of HIV PR complexed with TL-3 (PDB: 3TLH).

Fig. 2 Standard nomenclature P**1** __ P*n*, P**1**__ P*n* used to designate amino acid residues of peptide substrates. The corresponding binding sites on the protease are referred to as S**1** _ Sn, S**1**__ S*n* subsites.

Fig.3: TL-3inhibitor





2.5 Active site

Highly conserved catalytic sequence: *Asp-Thr-Gly* (25-26-27) in each monomer [8].*Thr* is buried in active site. *Asp* is essential to PR both catalytically and structurally *Asp*25 (25') induces a general acid/ base protein hydrolysis. The dimeric structure, in which each monomer contributes one *Asp-Thr-Gly* triad to the pseudo-symmetric active site, shows an active site that is indistinguishable from those of the monomer *Asp*artic proteases.However,the results from most of these studies are consistent with general acid–base mechanism, in which the two active site *Asp*artate residues play an essential general acid–base role to activate the water molecule that acts as a nucleophile and attacks the carbonyl carbon of the scissile bond. It was generally believed that this water molecule is located between the active site *Asp*artates, although some have suggested a different nucleophilic water molecule.

2.6 Secondary structure

HIV PR is a 99 amino acid *Asp*artyl protease, which functions as a homodimer with only one active site, which is *C*2-symmetric in the free form. More than 140 structures of the HIV-1 PR, its Each monomer contains an extended β-sheet region (a *Gly*cine-rich loop) known as the flap, that constitutes in part the substrate-binding site and plays an important role in substrate binding, and one of the two essential *Asp*artyl residues, *Asp*-25 and *Asp*-25' which lie on the bottom of the cavity.

Fig.4: One monomer of HIV-1 PR **Fig.5:** Dimer of HIV-1PR

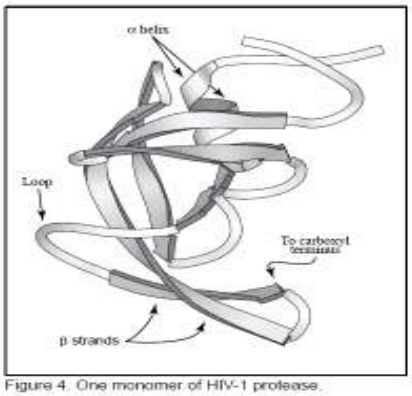


Figure 4. One monomer of HIV-1 protease.

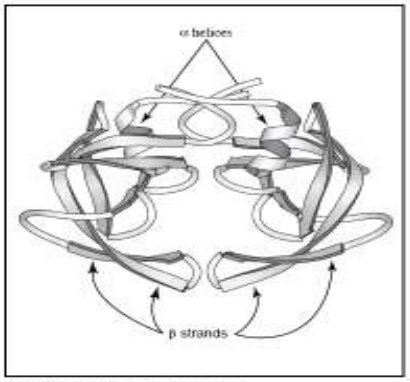


Figure 5. HIV-1 protease dimer.

2.7 Mechanism of the HIV protease

Proteases are known to play essential roles in many biological processes. They catalyze the hydrolysis of peptide bonds with high sequence selectivity and catalytic proficiency. These enzymes accomplish their catalysis by two different mechanisms, thus dividing them mechanistically into two broad classes of protease enzymes. The first class of enzymes uses an activated water molecule to attack the amide bond carbonyl of the substrate's scissile bond. The activation of the water molecule can be achieved either by a zinc cation (the zinc metalloproteinase's) or by the two-*Asp*artyl β-carboxy groups at the active site (the *Asp*artate proteases). In the second class of proteases, a nucleophilic atom of an amino acid side chain is used to initiate amide hydrolysis. In the first step the nucleophilicatom, which can be a hydroxyl group or a thiol, is activated by another amino acid side chain. The activated nucleophile attacks the carbonyl of the scissile amide bond to form an ester or a thioester acyl intermediate. Eventually, a water molecule to the corresponding hydrolysis products hydrolyzes this acyl enzyme intermediate. According to several studies, HIV PR in general has been shown to belong to the class of the *Asp*artic proteases. Examining the sequence homology of HIV PR to other cellular *Asp*artic proteases shows that this enzyme has the sequence *Asp-Thr-Gly*, which is conserved among the *Asp*artic protease enzymes [8]. These results suggested that HIV PR-1 enzyme might have similar structural features to the *Asp*artic protease enzymes as well as a similar mechanism. Indeed, mutational analysis by several groups of the highly conserved *Asp*- 25 has shown that substituting this

residue with *Asn* [2, 9], *Thr* [10] or Ala [11**]** leads to a protein without any proteolytic activity. More support for HIV PR being a member of the *Asp*artic protease family came from the *in vitro* inhibition of this enzyme by pepstatin, a natural product that selectively inhibits members of this family [9, 10, 12]. Finally, the *Thr*ee-dimensional structure of this enzyme also supported the classification of HIV PR in the *Asp*artic protease family [4, 5, 13]**.**The dimeric structure, in which each monomer contributes one *Asp-Thr-Gly* triad to the pseudo-symmetric active site, shows an active site that is indistinguishable from those of the monomeric *Asp*artic proteases. Kinetic methods, affinity labeling and X-ray crystallography have extensively studied the catalytic mechanism of the non-viral *Asp*artic proteases. Several mechanisms of action for this family proposed, including a mechanism that involves the formation of a covalent intermediate.

2.7.1 Catalytic mechanism for HIV

Fig.2 Standard nomenclature P**1** __ P*n*, P**1**__ P*n* is used to designate amino acid residues of peptide substrates. The corresponding binding sites on the protease are referred to as S**1** _ S*n*, S**1**_ S*n*_ subsites. general acid–base mechanism, in which the two active site *Asp*artate residues play an essential general acid–base role to activate the water molecule that acts as a nucleophile and attacks the carbonyl carbon of the scissile bond. It is generally believed that this water molecule is located between the active site *Asp*artates, although some have suggested a different nucleophilic water molecule [15]**.** The most widely accepted mechanism for *Asp*artic protease has been described by Suguna *et al.* (Fig. 3)[16]**.** The proposed mechanism is based on the crystal structure of the *Rhizopys chinensis Asp*artic protease complexed with a reduced peptide inhibitor. The pH–rate profile of this enzyme implies that only one of the two active site *Asp*artic acids is unprotonated in the active pH range [17]. In the proposed mechanism, the *Asp* group that is closer to the nucleophilic water molecule was assigned the negative charge (Fig.3). The nucleophilic water molecule held between the catalytic *Asp*artates, after its activation by the negative *Asp*artate side chain, attacks the carbonyl group in the substrate scissile bond to generate an oxyanion tetrahedral intermediate. Protonation of the scissile amide N atom and rearrangement result in the breakdown of the tetrahedral intermediate to the hydrolysis products. This general acid–base mechanism of the *Asp*artic protease family precludes the use of a Lewis acid

such as Zn^{+2}(as in the case of zinc metalloproteinase) and the formation of covalent acyl intermediates [14, 18]**.** Although the HIV PR mechanism shares many features with the rest of the *Asp*artic protease family, the full detailed mechanism of this enzyme remains not fully understood. Jaskólski *et al.* have proposed a new model of the enzymatic mechanism for the HIV PR enzyme based on the crystal structure of a complex between a chemically synthesized HIV PR and an octapeptide inhibitor U-85548e (H-Val-Ser-Gln-Asn-Leu-ψ-[CH(OH)- CH**2**]-Val-Ile-Val-OH)[19**].** As well as other crystal structures of HIV PR complexed with different inhibitors [20**]** in this mechanism (Fig.6). The hydrolysis reaction is viewed as a one-step process during which the nucleophilic water molecule and the acidic proton attack the scissile peptide bond in a concerted manner. The issue of the simultaneous attack from the nucleophile and electrophile is the major difference between this mechanism and other previously proposed mechanisms. As in the case of all the *Asp*artic protease family, the possibility of covalent catalysis (*e.g. Asp*-25 attacks directly the carbonyl amide bond) in the chemical mechanism of HIV PR is discounted. Indeed, Hyland *et al.* have provided evidence against the formation of a covalent intermediate by studying

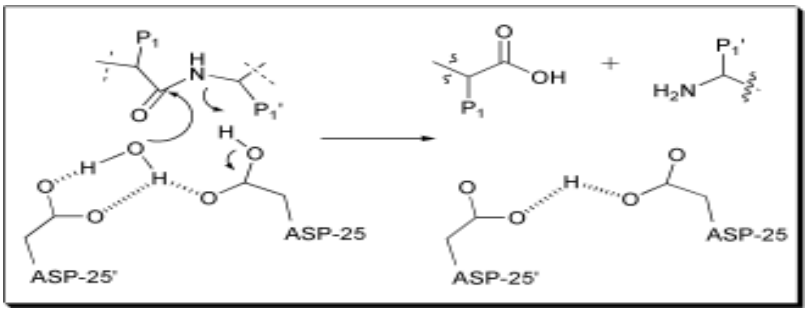


Fig.6: Proposed concerted catalytic mechanism for HIV.

It has been suggested that the protonation state of the two-*Asp*artic groups depends on the local environment near the *Asp*artate and is different for different inhibitors [19, 23]. However, it has been suggested also that the difference in p*K***a**s of these residues is more a function of their proximity to each other than of their differing environments [13, 24**].** *Asp*artic proteases are unique in that they function physiologically over a wide pH range (2–7.4)[17]. The pH–rate profile of a model substrate for HIV PR studied by Meek and coworker shows that the two *Asp*artate groups have different p*K*a values of 3.1 and 5.2[25]**.**

In contrast to these results, Smith *et al.* have found in their NMR experiments on the 13C-enriched enzyme in the absence of inhibitor, that the two *Asp* side chains are Fig.6 Proposed concerted catalytic mechanism for HIV. Equivalent and are both deprotonated at pH 6[26]. However, in the presence of pepstatin inhibitor only one *Asp* side chain is deprotonated in the pH range 2.5–6.5. Wang *et al.* have also examined the protonation state of the two-*Asp*artic groups, however using the asymmetric inhibitor KNI-272[27]. They also found one catalytic *Asp* to be protonated and the other unprotonated. In contrast to these results, Yamazaki *et al.* have Studied the ionization state of the two-*Asp*artic groups in HIV PR complexed with the symmetric inhibitor DM323[28]. This non-peptidic inhibitor contains two hydroxyl groups and forms a completely symmetric complex with the enzyme, in which the side chains of the two *Asp*artates were found to be magnetically equivalent and both protonated over the pH range 2–7. Using *ab initio* molecular dynamics studies of the pepstatin A–HIV PR complex, Piana *et al.* have proposed recently that both *Asp*artic groups are protonated [29] in contrast to what had been reported by Smith *et al* [26]. An answer to the question of where the acidic proton is located in the free enzyme could not be easily provided since proton locations are generally not resolvable by X-ray crystallography methods [16, 19] Meek *et al.* have shown that the known *Asp*artic protease inactivate 1, 2-epoxy-3-(4-nitrophenoxy) propane produced irreversible, time-dependent inactivation of HIV PR, *Thr*ough covalent modification of the enzyme's *Asp*artyl residue [30]. The PH-dependent kinetics of this inhibitor–enzyme interaction was consistent with having one protonated *Asp*artic group in the active site of the enzyme. Following this observation, Lee and Coworkers [31] have designed several inhibitors based on a peptide isostere containing *cis*-epoxide for the irreversible inactivation of HIV.These results are surprising, since one would expect the catalytic water molecule to be involved in the epoxide ring opening instead of direct attack by the *Asp*artyl group. Perhaps the alignment of the inhibitor in the active site does not allow enough space between the epoxide ring and the two-*Asp*artyl residues to accommodate a water molecule. Hyland *et al.* have proposed a detailed chemical mechanism for HIV PR [32] based on kinetic data obtained from solvent kinetic isotope effects, pH–rate and O incorporation studies [21, 25]. Combined with known previous structural data of HIV PR. In this mechanism, *Asp*-25 exists in the unprotonated state (p*K***a** 3.1) upon binding to substrate, while the proton on *Asp*-25 (p*K***a**

5.2) is positioned to hydrogen bond to the carbonyl oxygen of the substrate, and at the same time, the lytic water is positioned closer to the β-carboxylic acid of *Asp*-25_. This mechanism has many similar features in common with the general acid–base mechanism of *Asp*artic proteases reported by other groups [15, 16, 33]. It also resembles in part the concerted mechanism proposed by Jaskólski [19] in that the amine product is protonated by an active site *Asp*artyl residue.

Chapter 3

OBJECTIVE OF THE WORK

3. OBJECTIVE OF THE WORK

The study undertaken in this thesis was part of an ongoing research project,

This has overall aim of identifying new approach to design HIV-1 Protease inhibitors.

The Specific objectives of the thesis were:

- Prediction of binding energies by various molecular docking methods.
- Comparison of predicted binding energies to experimental binding energies.
- Identifying the good molecular docking method among various docking methods.
- Consensus ranking of enzyme-inhibitor binding in various molecular docking methods.

Chapter 4

MATERIALS AND METHODS

4.1 Protein Data Bank (PDB)

4.2 UCSF Chimera

4.3 Various molecular docking methods

4.4 AutoDock 4.0

4.5 ParDock

4.6 PatchDock

4.7 FireDock

4.1. Protein Data Bank (PDB):

All complexes deposited at PDB with a resolution and R-factor less than 3.0 and 0.2 respectively, were taken for the study (http://www.rcsb.org/pdb/home/home.do). Twenty-five HIV-1 protease-inhibitor complexes were taken from Protein Data Bank in the form of PDB format.

Table1: For each complex, the PDB code, resolution, R-factor.

S.NO	PDB	Resolution(A^0)	R-factor
1	1a9m	2.30	0.185
2	1aaq	2.50	0.190
3	1ajv	2.0	0.187
4	1ajx	2.0	0.161
5	1gno	2.30	0.17
6	1g2k	1.95	0.200
7	1hbv	2.30	0.18
8	1hih	2.30	0.19
9	1hiv	2.0	0.17
10	1hpo	2.50	0.179
11	1hpv	1.90	0.19
12	1hpx	2.0	0.170
13	1hsg	2.0	0.166
14	1hte	2.80	0.16
15	1htf	2.20	0.19
16	1htg	2.00	0.19
17	1hvh	1.80	0.88
18	1hvj	2.0	0.158
19	1hvr	1.80	0.19
20	1hxw	1.80	0.201
21	1ida	1.70	0.196

22	1qbr	1.80	0.190
23	1qbt	2.10	0.190
24	2upj	3.0	0.14
25	9hvp	2.80	0.18

The training set used in this study comprises 25 protein ligand complexes (Table 1). All the complexes were taken from the Protein Data Bank (PDB) [35]. Since our interests are concentrated on small non-covalently bound ligands, those complexes containing covalently bound ligands, complex ligands and macromolecular ligands were stripped out of the data set. More than seventy different kinds of proteins are represented in this training set and all the structures are of high resolution (better than 3.0 Å). The experimentally determined binding data were cited from the literature [35, 36] and expressed in the binding energies (table.2).

Table.2: Twenty-five HIV-1 Protease-inhibitor binding energies by experimental determined and predicted values with RMSD values.

S.NO	PDB	Experimental Binding Energy (kcal/mol)
cvcvc 1	1a9m	-9.41
2	1aaq	-11.45
3	1ajv	-10.59
4	1ajx	-10.86
5	1gno	-10.57
6	1g2k	-10.81
7	1hbv	-8.68
8	1hih	-10.97
9	1hiv	-12.64
10	1hpo	-11.82
11	1hpv	-12.58
12	1hpx	-12.53
13	1hsg	-12.93
14	1hte	-7.69
15	1htf	-11.04
16	1htg	-13.21
17	1hvh	-10.81
18	1hvj	-14.26
19	1hvr	-12.97
20	1hxw	-14.54
21	1ida	-11.86
22	1qbr	-14.42
23	1qbt	-14.49
24	2upj	-10.14
25	9hvp	-11.38

4.2. Chimera:

UCSF Chimera is a highly extensible, interactive molecular visualization and analysis system. Chimera can read molecular structures and associated data in a large number of formats, display the structures in a variety of representations, and generate high-quality images and animations suitable for publication and presentation. In addition, Chimera provides tools to: show density maps and analyze microscopy data; utilize symmetry information for the display of higher-order structures; display multiple sequence alignments, with crosstalk between the sequences and structures; and enable analysis of molecular dynamics trajectories and docking results.

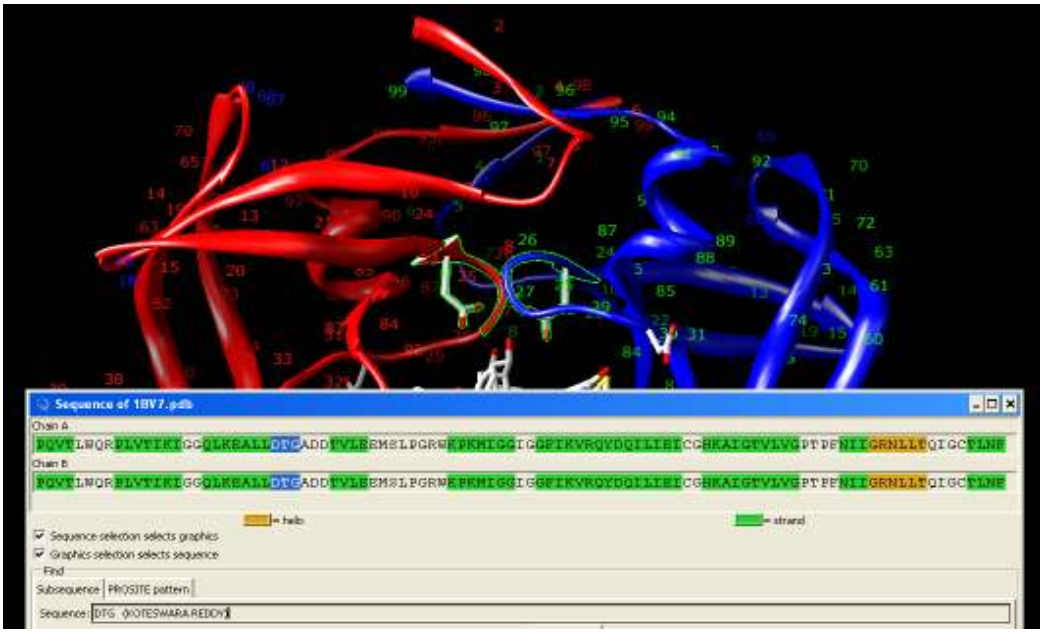
Molecular Graphics

- interactively manipulable stick, ball-and-stick, CPK, ribbon, and special nucleotide representations; molecular surfaces
- highly intuitive translation, scaling, and rotation; Side View tool for adjusting clipping planes and scaling
- interactive color editing in various color spaces (RGB, CMYK, etc.), including transparency
- ability to save high-resolution images for presentation and publication
- stereo viewing (side-by-side and time-sequential)

Chemical Knowledge

- determination of atom types in arbitrary molecules, including non-standard residues
- ability to add hydrogen atoms
- high-quality hydrogen bond identification
- selection of atoms/bonds by element, atom type, functional group, amino acid category
- interactive bond rotation, distance and angle measurements

Fig.7: *Asp*artic HIV-1 protease Active site: *ASP*25-*THR*26-*GLY*27 shows in chimera 3D protease strcture.



4.3. Various molecular docking methods

Given recent improvements in search algorithms and energy functions, computational docking methods have become a valuable tool to probe the interaction between an enzyme and its inhibitors. The interaction energy between the protein and its ligand is calculated by a simplified, often grid-based force field [37, 38]. Basic components may include steric and electrostatic energies, sometimes supplemented by other terms accounting for hydrogen bonding and salvation effects.

These methods can contribute significantly to the understanding of structural and energetic basis of enzyme-substrate interactions. Protein-ligand docking methods aim to predict the binding energy of the protein-ligand complex given the atomic coordinates. In such calculations, both the protein and ligand can be treated as rigid bodies [39]. We performed protein rigid and ligand flexible docking by autodock4.0, pardock, patchdock and firedock methods. Pardock, patchdock and firedock are online docking methods. To judge the performance of any approach for prediction of binding energy and consensus ranking, it is

important to consider the various types of docking methods and scoring functions used. Computationally top-scoring solutions taken a selection from among the top scoring solution (e.g. the one closest to the experimental data among the ten top solutions) or simply the solution closest to the experimental data[40] (easiest and least rigorous approach). The accurate prediction of enzyme-substrate interaction energies is one of the major challenges in computational biology. Generally various docking methods followed by various energy scoring functions.

We performed various Molecular Docking simulations on HIV-1 protease-inhibitor complexes and using the resulting structures to calculate the binding energies by AutoDock 4.0 and online molecular docking tools pardock, patchdock and firedock programs. Autodock and pardock produced binding free energy of enzyme-inhibitor and global energy produced by firedock, which is binding free energy of solution.Complementarity score produced by patchdock. Common top ranked docking solutions in autodock and pardock showed that good binding in molecular interaction. Hence consensus scoring or ranking of top ten solutions become a good binding in molecular interactions in these methods. The consensus ranking of enzyme-inhibitor complexes in various molecular docking methods improve the binding energy predictions. Consensus ranking has become an important method in various molecular docking methods to identify new inhibitors in computer-assisted discovery of new pharmaceutics.

4.4. Autodock 4.0

AutoDock is a suite of automated docking tools. It was designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. AutoDock actually consists of two main programs: AutoDock performs the docking of the ligand to a set of grids describing the target protein [41]; Auto Grid pre-calculates these grids. In additions to using them for docking, the atomic affinity grids can be visualized. This can help, for example, to guide organic synthetic chemists design better binders. autodock tools, or ADT developed a graphical user interface called for short, which amongst other things helps to set up which bonds will treated as rotatable in the ligand and to analyze dockings.

4.4.1 AutoDock has applications in

- X-ray crystallography
- structure-based drug design
- lead optimization
- virtual screening (HTS)
- combinatorial library design
- protein-protein docking
- Chemical mechanism studies.

AutoDock is used to perform computational molecular docking of small molecules to proteins, DNA, RNA and other important macromolecules, by treating the ligand and selected parts of the target as conformationally flexible. It uses a scoring function based on the AMBER force field, and estimates the free energy of binding of a ligand to its target. Novel hybrid global-local evolutionary algorithms are used to search the phase space of the ligand-macromolecule system.

The introduction of AutoDock 4 comprises three major improvements:

1. The docking results are more accurate and reliable.
2. It can optionally model flexibility in the target macromolecule.
3. It enables AutoDock's use in evaluating protein-protein interactions.

AutoDock 4.0 not only is it faster than earlier versions, it allows side chains in the macromolecule to be flexible. As before, rigid docking is blinding *Gly* fast, and high-quality flexible docking can be done in around a minute. Up to 40,000 rigid dockings can be done in a day on one CPU.AutoDock 4.0 now has a free energy scoring function that is based on a linear regression analysis, the AMBER force field, and an even larger set of diverse protein-ligand complexes with known inhibition constants than we used in AutoDock 3.0. The best model was cross validated with a separate set of HIV-1 protease complexes, and confirmed that the standard error is around 2.5 kcal/mol. This is enough to discriminate between leads with milli-, micro- and nano-molar inhibition constants.

AutoDock 4.0 can be compiled to take advantage of new search methods from the optimization library, ACRO, developed by William E. Hart at Sandia National Labs. We

have also added some new features to our existing evolutionary methods. We still provide the Monte Carlo simulated annealing (SA) method of 2.4 and earlier. The Lamarckian Genetic Algorithm (LGA) is a big improvement on the Genetic Algorithm, and both genetic methods are much more efficient and robust than SA.

4.4.2 AutoDockTools (ADT)

We have developed and continue to improve our graphical front-end for AutoDock and AutoGrid, ADT (AutoDockTools). It runs on Linux, Mac OS X, SGI IRIX and Microsoft Windows. AutoGrid is a program that pre-calculates grid maps of interaction energies for various atom types, such as aliphatic carbons, aromatic carbons, hydrogen-bonding oxygen, and so on, with a macromolecule such as a protein, DNA or RNA. These grid maps were then used by autodock docking calculations to determine the total interaction energy for a ligand with a macromolecule. Doing this pre-calculation saves a lot of time during the docking, primarily because we do not have to update non-bonded lists during the calculation. In addition, what was a calculation with order N-squared complexity is reduced to one that is order N, where N is the number of atoms interacting.

AutoDock 4 offers many new features and improvements over previous versions. The most significant is that it models flexible side chains in the protein. We have recalibrated the free energy scoring function using a much larger training set of ligand-protein complexes. We can get both the 3D structure and the inhibition constants, Ki. There are also new search methods.

Many limitations of version 3 have been removed in AutoDock 4. One of biggest annoyances was the limit of 6 different atom types in a ligand; this has been increased to 22, and the method of specifying atom types is improved, using a single parameter file called "AD4_parameters.dat" by default. In addition, autodock 4 no longer requires the user to set the stack size to unlimited in the UNIX shell.

4.4.3 Importance of Autodock

It is very fast, provides high quality predictions of ligand conformations, and good correlations between predicted inhibition constants and experimental ones. AutoDock has also been shown to be useful in blind docking where the location of the binding site is not

known. Autodock4.0 introduces the ability to model conformational flexibility in selected side chains in the target macromolecule; it uses a new desolvation energy term and was calibrated on a set of 188 ligand-protein complexes of known structure and binding affinity. There have been numerous improvements made to the search algorithms and the molecular representation of the problem, and autodock supports more atom types than previous versions. Thanks to ADT, a very refined graphical user interface, autodock has been made easier to use than earlier versions, allowing bench chemists and other non-computational chemists to set up dockings and analyse the results visually.

4.4.4 AutoDock4 scoring functions

- van der Waals
- Hydrogen Bonding
- Electrostatics
- Desolvation
- Torsional

Binding energy=Intermolecular energy+Torsional energy

$\Delta G_{bind} = \Delta G_{vdw} + \Delta G_{ele.} + \Delta G_{H\text{-}bond} + \Delta G$ desolv $+\Delta G_{tors}$

Here ΔG=change in free energy

AutoDockTools, or ADT, is the ultimate GUI to set up, launch and analyze AutoDockruns**.** View molecules in 3D, rotate & scale in real time.

- Add all hydrogen's or just non-polar hydrogen's.
- Assign partial atomic charges to the ligand and the macromolecule (Gasteiger or Kollman United Atom charges).
- Merge non-polar hydrogen's and their charges with their parent carbon atom.
- Set up rotatable bonds in the ligand using a graphical version of AutoTors.
- Set up the AutoGrid Parameter File (GPF) using a visual representation of the grid box, and slider-based widgets.
- Set up the AutoDock Parameter File (DPF) using forms.
- Launch AutoGrid and AutoDock.
- Read in the results of an AutoDock job and graphically display them.
- View is contoured AutoGrid affinity maps.

4.5. Pardock

This program all-atom energy based Monte Carlo docking procedure tested on a dataset of 226 protein-ligand complexes [36]. Average root mean square deviation (RMSD) from crystal conformation was observed to be ~ 0.53 Å. The correlation coefficient (r^2) for the predicted binding free energies calculated using the docked structures against experimental binding affinities was 0.72. The docking protocol is web-enabled as free software at (www.scfbioiitd.res.in/dock) ParDOCK will predict the binding mode of the ligand conforming to the following format and other requirements:

4.5.1 Online procedure:

1. Upload hydrogen added Protein Reference Complex and Candidate Molecule in PDB format.
2. Please make sure that Protein Reference Complex contains the Drug information to perform docking.
3. Enter Formal Charge between +10 to -10. If no charge is entered, '0' will be assigned by default.
4. Specify your email Id in text box. As the program might take 15-20 minutes to process the query, result will be e-mailed to you. If you do not specify any email id, Job ID that is provided at each submission can retrieve result. The job ID can also be used to check the status of your job.
5. Click Run to submit the job. If other jobs are running on the server, your job will be put in queue and Job Status will be shown to you.

4.5.2 PDB format

the input file should be in the 'AMBER' PDB format i.e., as described here

1. There is a slight difference in the format for Protein Reference Complex and drug, as per PDB convention, this is shown below.
2. The spacing between columns is not important but two columns should NOT merge at any place.

The structural inputs for ParDOCK are a reference complex (target protein bound to a reference ligand) and a candidate molecule. Fig.8 shows the flowchart of docking methodology adopted. The ParDOCK protocol consists of four main steps: (a)

identification of the best possible grid/ translational points in radius of 3Å around the reference point (centre of mass);(b) generation of protein grid and preparation of energy grid in and around the active site of the protein to pre-calculate the energy of each atom in the candidate ligand; (c) Monte Carlo docking and intensive configurational search of the ligand inside the active site; and (d) identification of the best docked structures on an energy criterion and prediction of the binding free energy of the complex.

This completely automated version of ParDOCK was developed for finding the binding mode of the ligand to its receptor to a known binding site and not for predicting all possible binding sites. The reference complex therefore helps in initiating the search. For the sake of efficiency, a portion of the receptor enclosing the binding site is considered and this simplification was accounted for in atomic level energy calculations. Goodford pioneered the use of grids for docking protocols and we have used grid representations for energy calculations. The basic idea is to store information about the receptor's energy contributions on grid points so that only these need to be read during scoring.

4.5.3 Methodology

1. Identification of the Grid / Translation Points

The center of mass (COM) of the reference ligand is calculated and considered as a reference point inside the active site of the protein. A cube of length 6Å (i.e. ± 3Å) with respect to the reference point is created and a uniform grid of length 1Å inside the 6Å cube is defined. The 1Å grids occupied by protein atoms are eliminated and the free grid points are considered. This helps to increase the efficiency of search by providing all possible translational points in a decreased/ vacant search space.

2. Generation of Protein Grid and Pre-Calculation of Energy

Protein grid of length 10Å is generated in order to identify the protein residues interacting with the candidate ligand. The protein grid formation helps in identifying the protein residues in a specified range, which are interacting with atoms of the candidate ligand. While searching for the spatial positions around the reference point (COM) within the active site of the protein, the clash module helps in identifying the appropriate translation points from the given series of points in the range of 1Å. Number of clashes is calculated at

each translation point and the best set of translation points with minimum number of clashes are selected in the given cube of length 6Å.

The energy grid was arranged in and around the active site of the protein based on the selected translation points and the energy points were placed at equal distance inside the grid. The interaction energy of each atom of the candidate ligand was pre-calculated at each point using a scoring function comprising three components: electrostatic, van der Waals and hydrophobic interactions. For a given energy point, a cube cut off is fixed and the protein atoms interacting with this point are identified.

3. Generation of Monte Carlo Configurations of the Ligand

The Monte Carlo configurations are generated around six degrees of freedom, which result in many combinations of ligand configurations at each translation point. For each Monte Carlo configuration, energy points were selected and pre-calculated energies of each atom are added. The putative binding configurations were selected based on the scoring function.

4. Preparation of Protein and Ligand for Docking

The complexes chosen for study are adapted from RCSB and prepared in a force field compatible manner. Initially the crystallographic water molecules were removed and the ligand coordinates are extracted from the protein-ligand complex. Hydrogen atoms were added keeping the ionization states of the atoms in the ligand as specified in the literature. The ligand is then geometry optimized *Thr*ough AM1 procedure followed by calculation of partial charges of the ligand by AM1-BCC procedure. GAFF force field was used to assign atom types, bond angle, dihedral and van der Waals parameters for the ligand.

5. Description of All-Atom Energy Based Scoring Function

The ligand configurations generated were ranked based on an all atom energy function, which calculates non-bonded interactions of protein-ligand complexes as described in equation (I).

$$E = \sum E_{el} + E_{vdw} + E_{hpb} \quad (I)$$

E is the total non-bonded energy, Eel is the electrostatic contribution to the energy, Evdw is the van der Waals term, Ehpb is the hydrophobic term and the summation runs over all the atoms of the protein-ligand complex. Electrostatic contribution is calculated by Coulomb's law with sigmoidal dielectric function, van der Waals term is computed using a (12, 6)

Lennard-Jones potential between the atoms of protein and ligand and hydrophobic interactions are calculated by Gurney parameter approach.

6. Energy Minimization of Docked Structures and Protein- Ligand Binding Free Energy Estimations

The selected docked complexes are energy minimized in vacuum by AMBER. For vacuum minimizations, 1000 steps of steepest descent and 1500 steps of conjugate gradient was carried out. The minimization procedure was repeated using explicit solvent, without much difference in the calculated energetics. Hence, the vacuum minimization protocol was retained due to its expeditious nature. The energy-minimized structure was employed in computing the binding affinity by a scoring function, BAPPL. The energy function employed in BAPPL includes contributions of electrostatics, van der Waals, hydrophobicity and loss of conformational entropy of protein side chains upon ligand binding.

Fig.8. Pardock methodology

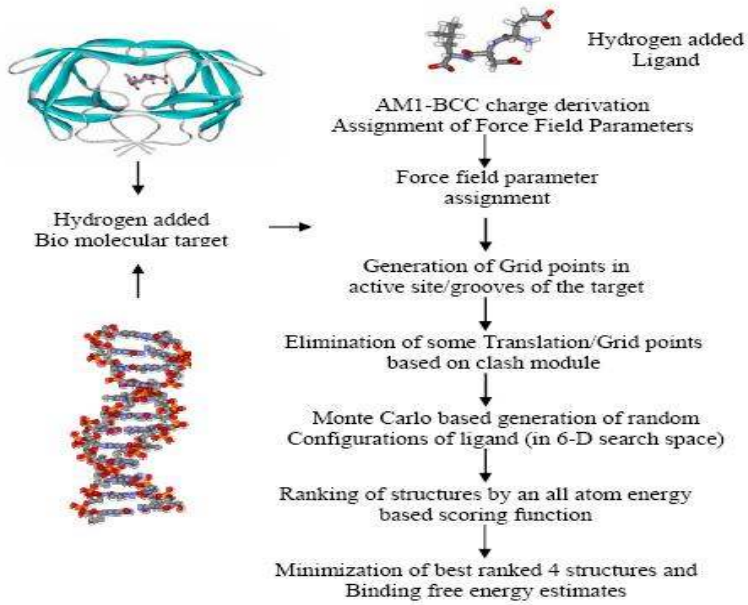


Figure: A computational flowchart for docking methodology

4.5.4 Prdock procedure:

1. Center of mass (COM) the reference ligand is calculated
2. The grid is made in the radii of $6A^0$ around the COM of the reference ligand
3. A grid length of $1A^0$ box inside $6A^0$ cube was made and each center point of $1A^0$ box was considered as a translation point.
4. Selected translation points were shown after eliminating some of the points using clash module.
5. Grid generated for the whole protein by $10A^0$ length
6. Based on selected translation point's energy grid was made at equal distance and pre energy calculation of each ligand atom at each energy point is calculated.
7. Generation of Monte Carlo configurations by six degrees of freedom
8. The energy of each configuration was matched with the energy points and the pre-calculated energies were added.
9. The best energy structure was selected based on energy points
10. The RMSD difference between crystal structure and energy minimized docked structure.

4.6. PatchDock

Patchdock [43] is a molecular docking algorithm based on shape complementarity principles. It aimed at finding docking transformations that yield good molecular shape complementarity. Such transformations, when applied, induce both wide interface areas and small amounts of steric clashes. A wide interface ensured to include several matched local features of the docked molecules that have complementary characteristics. The PatchDock algorithm divides the Connolly dot surface representation of the molecules into concave, convex and flat patches [44, 45]. Then, complementary patches matched in order to generate candidate transformations. Each candidate transformation further evaluated by a scoring function.

That considers both geometric fit and atomic desolvation energy [46]. Finally, an RMSD (root mean square deviation) clustering was applied to the candidate solutions to discard redundant solutions. The main reason behind PatchDock's high efficiency is its fast transformational search, which was driven by local feature matching rather than brute force searching of the six-dimensional transformation space. It further speeds up the computational processing time by utilizing advanced data structures and spatial pattern detection techniques, such as geometric hashing and pose clustering that were originally developed in the field of computer vision. The run time of PatchDock for two input proteins of average size (about 300 amino acids) is <10 min on a single 1.0 GHz PC processor under the Linux operating system.

4.6.1 The patchdock web server: input, output and user interface

It was developed a very simple and intuitive web interface for the PatchDock algorithm available at http://bioinfo3d.cs. tau.ac.il/PatchDock/. Once the docking request is submitted, the PatchDock algorithm starts the prediction process. The user is notified when the results are ready by an email message that contains a link to a web page where the predictions are presented. On this page, the user can both view specific predictions and download a compressed file of the top scoring solutions (see Figure 9).

Fig.9: patchdock input file

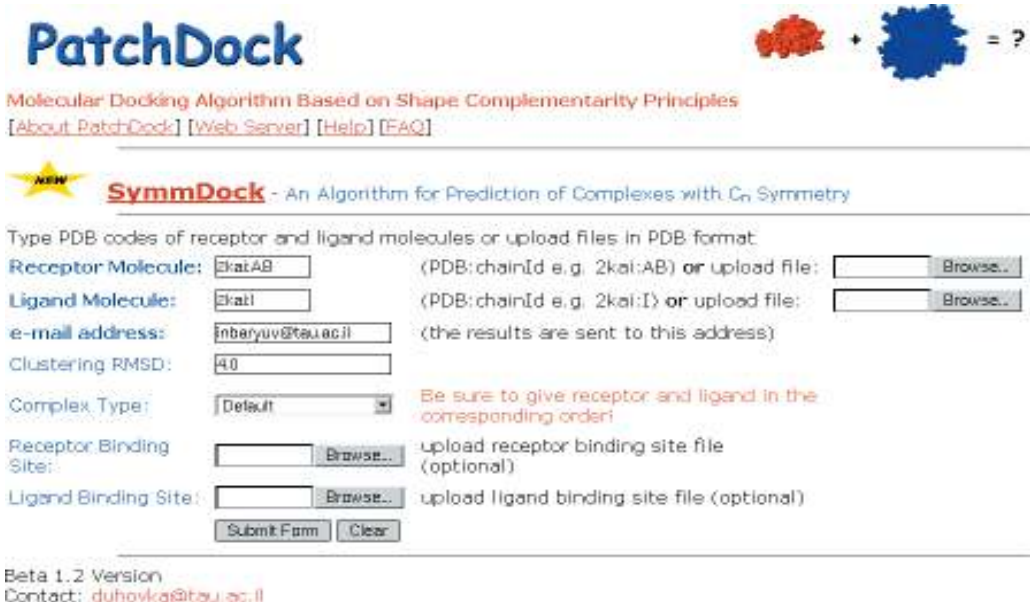
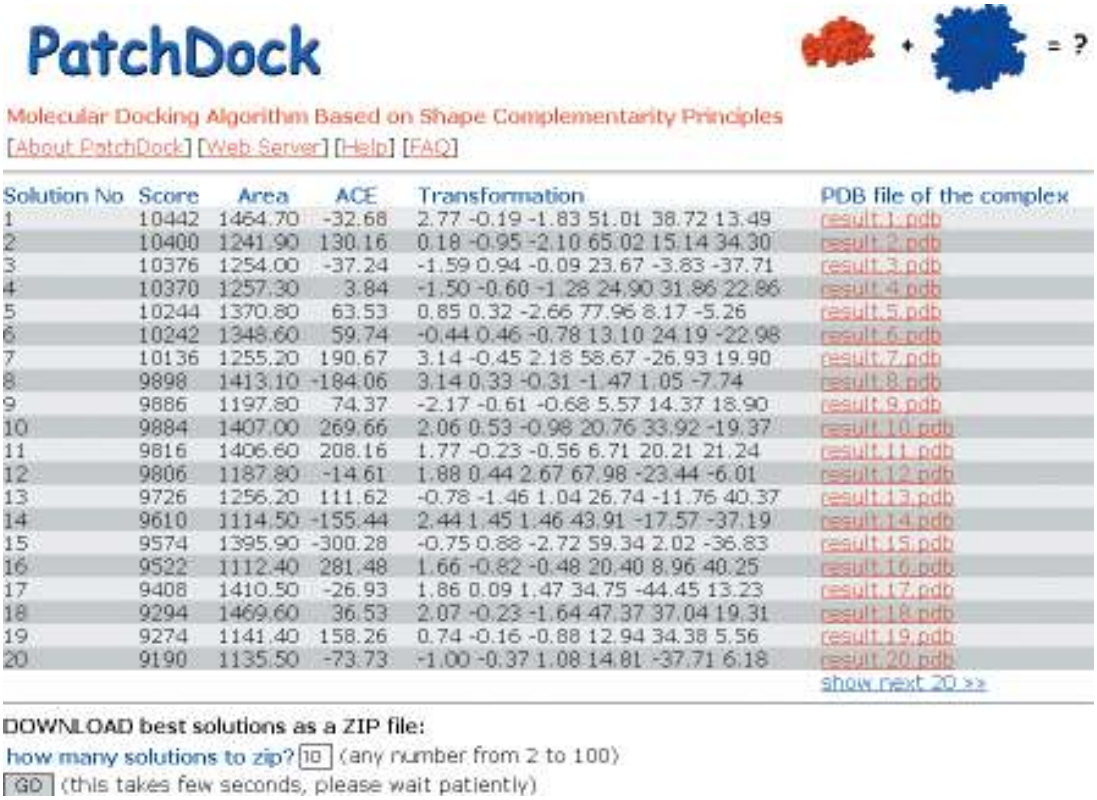


Fig.10: patchdock output file



PatchDock

Molecular Docking Algorithm Based on Shape Complementarity Principles

[About PatchDock] [Web Server] [Help] [FAQ]

Solution No	Score	Area	ACE	Transformation	PDB file of the complex
1	10442	1464.70	-32.68	2.77 -0.19 -1.83 51.01 38.72 13.49	result.1.pdb
2	10400	1241.90	130.16	0.18 -0.95 -2.10 65.02 15.14 34.30	result.2.pdb
3	10376	1254.00	-37.24	-1.59 0.94 -0.09 23.67 -3.83 -37.71	result.3.pdb
4	10370	1257.30	3.84	-1.50 -0.60 -1.28 24.90 31.86 22.86	result.4.pdb
5	10244	1370.80	63.53	0.85 0.32 -2.66 77.96 8.17 -5.26	result.5.pdb
6	10242	1348.60	59.74	-0.44 0.46 -0.78 13.10 24.19 -22.98	result.6.pdb
7	10136	1255.20	190.67	3.14 -0.45 2.18 58.67 -26.93 19.90	result.7.pdb
8	9898	1413.10	-184.06	3.14 0.33 -0.31 -1.47 1.05 -7.74	result.8.pdb
9	9886	1197.80	74.37	-2.17 -0.61 -0.68 5.57 14.37 18.90	result.9.pdb
10	9884	1407.00	269.66	2.06 0.53 -0.98 20.76 33.92 -19.37	result.10.pdb
11	9816	1406.60	208.16	1.77 -0.23 -0.56 6.71 20.21 21.24	result.11.pdb
12	9806	1187.80	-14.61	1.88 0.44 2.67 67.98 -23.44 -6.01	result.12.pdb
13	9726	1256.20	111.62	-0.78 -1.46 1.04 26.74 -11.76 40.37	result.13.pdb
14	9610	1114.50	-155.44	2.44 1.45 1.46 43.91 -17.57 -37.19	result.14.pdb
15	9574	1395.90	-300.28	-0.75 0.88 -2.72 59.34 2.02 -36.83	result.15.pdb
16	9522	1112.40	281.48	1.66 -0.82 -0.48 20.40 8.96 40.25	result.16.pdb
17	9408	1410.50	-26.93	1.86 0.09 1.47 34.75 -44.45 13.23	result.17.pdb
18	9294	1469.60	36.53	2.07 -0.23 -1.64 47.37 37.04 19.31	result.18.pdb
19	9274	1141.40	158.26	0.74 -0.16 -0.88 12.94 34.38 5.56	result.19.pdb
20	9190	1135.50	-73.73	-1.00 -0.37 1.08 14.81 -37.71 6.18	result.20.pdb

show next 20 >>

DOWNLOAD best solutions as a ZIP file:

how many solutions to zip? 10 (any number from 2 to 100)

GO (this takes few seconds, please wait patiently)

The PatchDock user interface (fig.9) the request form of PatchDock. The receptor molecule and the ligand molecule are given either by the PDB code of the molecule (chain IDs are optional) or by uploading a file in PDB format. (fig.10) the solutions page presents the geometric score, interface area size and desolvation energy of the 20 top scoring solutions. The user can use the 'show next 20' button to view solutions of lower score. The user can download each solution by pressing the solution link in the rightmost column or download an archive file (ZIP format) of the best solutions using the action button at the bottom of the page.

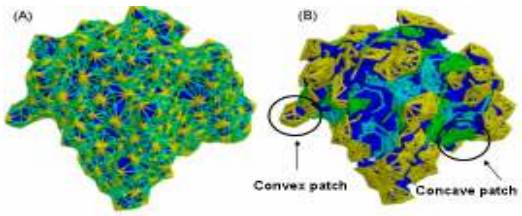
4.6.2 Molecular Docking Algorithm Based on Shape Complementarity Principles

PatchDock is an algorithm for molecular docking. The input is two molecules of any type: proteins, DNA, peptides, drugs. The output is a list of potential complexes sorted by shape complementarity criteria.

Short Overview: PatchDock algorithm was inspired by object recognition and image segmentation techniques used in Computer Vision. Docking was compared to assembling a jigsaw puzzle. When solving the puzzle we try to match two pieces by picking one piece and searching for the complementary one. PatchDock employs a similar technique. Given two molecules, their surfaces were divided into patches according to the surface shape. These patches correspond to patterns that visually distinguish between puzzle pieces. Once the patches were identified, they can be superimposed using shape matching algorithms. The algorithm has three major stages:

- **Molecular Shape Representation** - In this step, we compute the molecular surface of the molecule. Next, we apply a segmentation algorithm for detection of geometric patches (concave, convex and flat surface pieces). The patches were filtered, so that only patches with 'hot spot' residues were retained.

Fig.11: Molecular Shape Representation



- **Surface Patch Matching** - we apply a hybrid of the Geometric Hashing and Pose-Clustering matching techniques to match the patches detected in the previous step. Concave patches were matched with convex and flat patches with any type of patches.

- **Filtering and Scoring** - the candidate complexes from the previous step were examined. We discard all complexes with unacceptable penetrations of the atoms of the receptor to the atoms of the ligand. Finally, the remaining candidates were ranked according to a geometric shape complementarity score.

4.7. FireDock

The FireDock [47] server addresses the refinement problem of protein-protein docking solutions. The method simultaneously targets the problem of flexibility and scoring of solutions produced by fast rigid-body docking algorithms. Given a set of up to 25 potential

docking candidates, FireDock refines and scores them according to an energy function, spending about 3.5 seconds per candidate solution. To the best of our knowledge, this is the first web server that allows performing large-scale flexible refinement and scoring of docking solutions online.

4.7.1 Input files

There are two input options for FireDock. In the first option, the user uploads/specifies PDB codes of two PDB files (receptor and ligand) and provides a list of transformations. Each transformation, when applied on the ligand, produces a candidate docking solution. In the second option, the user can upload an input PDB file, with each docking solution represented by a MODEL. The candidate solutions for FireDock can be generated by rigid-body docking methods, such as Patchdock [48, 49], FFT-based methods such as ZDOCK, GRAMM-X, Hex, ClusPro etc. In addition, we provide an option for automatic redirection of solutions from PatchDock web server.

Each candidate is subsequently refined by restricted interface side-chain rearrangement and by soft rigid-body optimization. The side-chain flexibility was modeled by rotamers and integer linear programming [50] solves the obtained combinatorial optimization problem. Following rearrangement of the side-chains, the relative position of the docking partners is refined by Monte Carlo minimization of the binding score function. The binding score rank the refined candidates. This score includes Atomic Contact Energy [51], softened van der Waals interactions, partial electrostatics and additional estimations of the binding free energy.

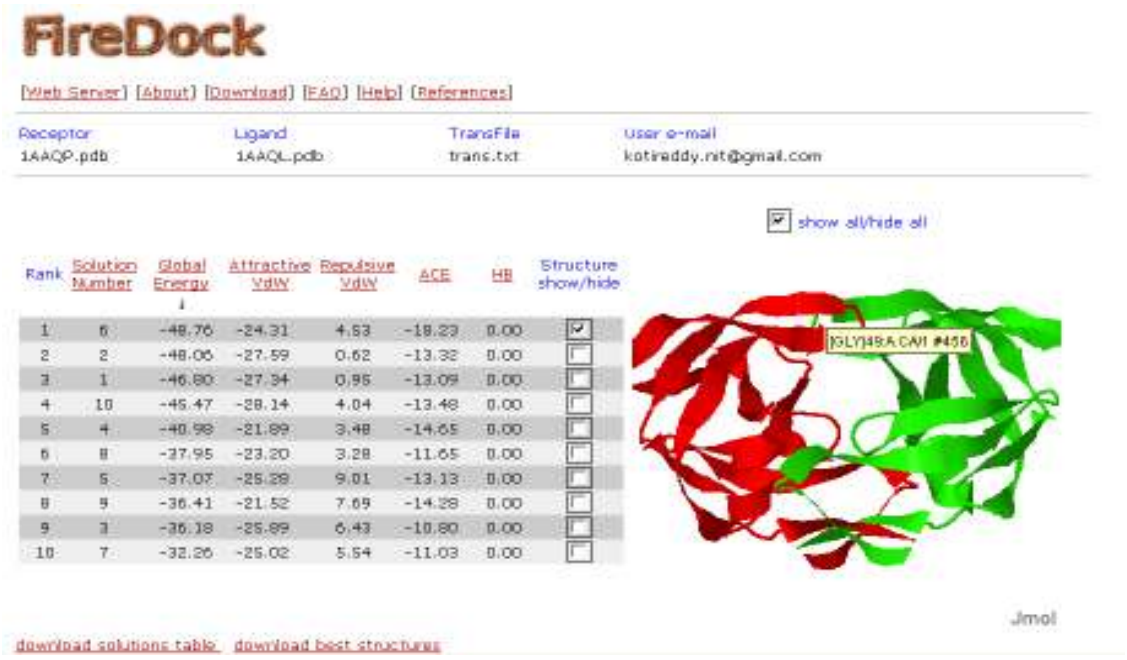
4.7.2 Output files

The output is a ranked list of all the input solutions. The refined complex structure is generated for up to 100 low-energy candidates. The user can view the complexes in the Jmol applet window and/or download the structures.

The figure below shows an example of FireDock usage. Transformations of docking candidates are generated by PatchDock [48, 49] and are given as an input to FireDock. First a coarse refinement is performed, using a restricted interface side-chain optimization with atomic radii scaling of 0.8, in order to allow a certain amount of steric clashes. The refined

candidates are scored and ranked according to the energy function and are returned as an output. Then, FireDock is run again on the best 25 solutions for a final refinement. In this second run, a full interface side-chain optimization is performed with atomic radii scaling of 0.85, in order to reduce the amount of clashes.

Fig.12: firedock output



4.7.3 Firedock scoring parameters:

Program parameters for score calculation for different complex types. Parameters are specified for the full and coarse refinement stages. The parameters described in the table are weights of the following scoring terms:

- attrV dWWeight - softened attractive van der Waals interactions
- repV dWWeight - softened repulsive van der Waals interactions
- attrElWeight - attractive short-range Coulomb electrostatics
- repElWeight - repulsive short-range Coulomb electrostatics
- l attrElWeight - attractive long-range Coulomb electrostatics
- l repElWeight - repulsive long-range Coulomb electrostatics
- ACEWeight - Atomic Contact Energy (ACE) potential

• HBWeight - hydrogen and disulfide Bonds
• catpiWeight - cation-_ stacking
• pipiWeight - _-_ stacking
• aliphWeight - aliphatic interactions
• insidenessWeight - “insideness” measure
• confProbWeight - internal energy

Chapter 5

RESULTS AND DISCUSSION

5.1. Binding energies predicted by Autodock 4.0

5.2. Binding energies predicted by Pardock

5.3. Patchdock and Firedock results

5.4. Comparison of binding energies in various methods

5.5. Correlation between experimental to predicted binding energy

5.1. Binding energy prediction by Autodock 4.0

Ligand all-atom RMSD (Å) and the number of docking solutions (*N*) in the cluster from 100 Lamarckian genetic algorithm (LGA) docking with lowest binding energy of each solution was top ranked solutions was selected(table.3). Moreover, Monte Carlo Docking Protocol runs for twenty-five protease-inhibitor complexes (Morris GM *et al.* 1998).

Table 3: For each complex, the lowest binding energy and RMSD evaluated by autodock.

S.NO	PDB	Autodock 4.0 Binding Energy (kcal/mol)	RMSD(A^0)	N
1	1a9m	-9.84	1.30	30
2	1aaq	-9.45	1.65	33
3	1ajv	-11.46	0.74	37
4	1ajx	-10.02	1.57	19
5	1gno	-12.19	1.73	6
6	1g2k	-7.53	0.00	39
7	1hbv	-9.04	1.92	96
8	1hih	-9.34	1.94	21
9	1hiv	-11.47	1.98	19
10	1hpo	-9.22	0.68	39
11	1hpv	-9.87	1.76	45
12	1hpx	-11.88	1.41	30
13	1hsg	-11.16	1.24	25
14	1hte	-7.54	1.95	22
15	1htf	-19.60	0.74	34
16	1htg	-13.29	1.22	80
17	1hvh	-8.03	0.01	5
18	1hvj	-10.03	1.80	48
19	1hvr	-14.80	0.13	16

20	1hxw	-10.04	0.82	49
21	1ida	-12.64	1.36	21
22	1qbr	-14.59	1.74	40
23	1qbt	-14.41	1.82	4
24	2upj	-10.20	1.46	13
25	9hvp	-11.43	1.71	53

Number of docking solutions in a cluster (N): Table.3 shows the number of docking solutions in a cluster (N) along with the all atom ligand RMSD for each docking solution. A small *N* value indicates strong specificity of binding, on average, the lower number of docking solutions in the cluster for all docking simulation indicates that the ligands bind to their binding pocket with high specificity, with all of the solutions resembling one of only a small number of binding conformations and orientations. On the other hand, if *N* is large, the experiment indicates a low specificity of binding, since the solutions are composed of many different binding conformations/orientations.

Average all-atom root mean square deviations (RMSD):

Table.3 shows the all atom RMSD between each simulated complex and the corresponding protease-inhibitor x-ray structure. The best results were obtained from the structures in pardock than autodock. These structures also have low (≅ 0.35 Å) average all-atom root mean square deviations (RMSD) relative to the experimental results (table.4). The average RMSD for the complexes docked using autodock4.0 were found to be ≅ 1.3 Å.

5.2. Binding energy prediction by Pardock

For each complex, the lowest binding energy and RMSD evaluated by pardock with Monte Carlo Docking Protocol runs for twenty-five protease-inhibitor complexes (Gupta *et al.* 2007).

Table.4: For each complex, the lowest binding energies and RMSD evaluated by pardock.

S.NO	PDB	ParDock	RMSD(A^0)
1	1a9m	-11.32	0.70
2	1aaq	-11.21	0.271
3	1ajv	-10.94	0.210
4	1ajx	-10.26	0.434
5	1gno	-10.10	0.290
6	1g2k	-12.13	0.370
7	1hbv	-10.91	0.442
8	1hih	-11.80	0.204
9	1hiv	-13.14	0.344
10	1hpo	-11.14	0.388
11	1hpv	-10.83	0.189
12	1hpx	-13.18	0.416
13	1hsg	-12.49	0.347
14	1hte	-8.06	0.219
15	1htf	-9.99	0.253
16	1htg	-12.86	0.440
17	1hvh	-12.18	0.454
18	1hvj	-15.02	0.206
19	1hvr	-12.96	0.195
20	1hxw	-14.60	0.327
21	1ida	-12.55	0.327
22	1qbr	-14.54	0.331
23	1qbt	-14.61	0.447
24	2upj	-10.70	0.493
25	9hvp	-13.90	0.515

5.3. Complementarily score and binding energy prediction by patchdock and firedock

The PatchDock algorithm divides the Connolly dot surface representation of the molecules into concave, convex and flat patches. Then, complementary patches were matched in order to generate candidate transformations. Each candidate transformation was further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy. Finally, an RMSD (root mean square deviation) clustering was applied to the candidate solutions to discard redundant solutions. The total complementarity score was produced in enzyme-inhibitor binding state. Top score solutions became the top ranked solutions were selected for each enzyme-inhibitor complex among 100 solutions (Table.5).

Firedock was produced global energy (binding energy) of enzyme-inhibitor binding state. Which is selected top ranked with lowest gobal energy of solution from top 10 solutions.which is followed by patchdock program.firedock was designed based on various scoring functions.we considered as global energy of each solution was divided by 5.55 factor which became a binding energy of each solution. The relation was generated between global energy and the binding energy in firedock solutions as follows

Firedock binding energy = firedock global energy/5.55

Table.5: Firedock and patchdock results

S.NO	PDB	Firedock Global energy	Firedock Binding energy *5.55	Patchdock Score
1	1a9m	-51.66	-9.31	8186
2	1aaq	-48.76	-8.78	8404
3	1ajv	-61.94	-11.16	8148
4	1ajx	-54.20	-9.76	8354
5	1gno	-87.37	-15.74	11084
6	1g2k	-62.16	-11.20	8462
7	1hbv	-64.97	-11.70	8874
8	1hih	-60.25	-10.85	8722
9	1hiv	-57.19	-10.30	10042
10	1hpo	-42.95	-7.74	6428
11	1hpv	-51.56	-9.29	7604
12	1hpx	-56.75	-10.22	8522
13	1hsg	-58.43	-10.53	8342
14	1hte	-38.38	-6.91	5912
15	1htf	-92.77	-16.71	10250
16	1htg	-132.58	-23.89	10002
17	1hvh	-52.39	-9.44	8014
18	1hvj	-59.72	-10.76	8836
19	1hvr	-67.03	-12.08	8854
20	1hxw	-67.64	-12.19	8716
21	1ida	-64.09	-11.55	9232
22	1qbr	-78.55	-14.15	9486
23	1qbt	-82.05	-14.78	9896
24	2upj	-52.67	-9.49	7864
25	9hvp	-74.88	-13.49	10292
Average:		-64.84	-11.68	8741

5.4. Comparison of binding energies

For each complex, the lowest binding energies evaluated by autodock, pardock, firedock and complementarities score predicted by patchdock programs. The differences between average experimental binding energies to predicted average binding energies in autodock, pardock and firedock were 0.55, 0.35 and 0.02. It was clearly illustrated that pardock binding energies are very near to experimental values with a difference 0.35. And firedock binding energy values are very near to experimental binding energy values with a difference 0.02 (if we considered as 5..55 multiple factor to predicted binding energy values). Patchdock was produced average score for all twenty-five complexes were 8741. The solutions with greater/equal to the average score were observed good binding solutions in enzyme-inhibitor binding process.

Table.6: comparison of binding energies

S.NO	PDB	Experimental Binding Energy (kcal/mol)	Autodock 4.0 Binding Energy (kcal/mol)	Pardock Binding Energy (kcal/mol)	Firedock Global Energy *5.55	Patchdock Score
1	1a9m	-9.41	-9.84	-11.32	-9.31	8186
2	1aaq	-11.45	-9.45	-11.21	-8.78	8404
3	1ajv	-10.59	-11.46	-10.94	-11.16	8148
4	1ajx	-10.86	-10.02	-10.26	-9.76	8354
5	1gno	-10.57	-12.19	-10.10	-15.74	11084
6	1g2k	-10.81	-7.53	-12.13	-11.20	8462
7	1hbv	-8.68	-9.04	-10.91	-11.70	8874
8	1hih	-10.97	-9.34	-11.80	-10.85	8722
9	1hiv	-12.64	-11.47	-13.14	-10.30	10042
10	1hpo	-11.82	-9.22	-11.14	-7.74	6428
11	1hpv	-12.58	-9.87	-10.83	-9.29	7604
12	1hpx	-12.53	-11.88	-13.18	-10.22	8522

13	1hsg	-12.93	-11.16	-12.49	-10.53	8342
14	1hte	-7.69	-7.54	-8.06	-6.91	5912
15	1htf	-11.04	-19.60	-9.99	-16.71	10250
16	1htg	-13.21	-13.29	-12.86	-23.89	10002
17	1hvh	-10.81	-8.03	-12.18	-9.44	8014
18	1hvj	-14.26	-10.03	-15.02	-10.76	8836
19	1hvr	-12.97	-14.80	-12.96	-12.08	8854
20	1hxw	-14.54	-10.04	-14.60	-12.19	8716
21	1ida	-11.86	-12.64	-12.55	-11.55	9232
22	1qbr	-14.42	-14.59	-14.54	-14.15	9486
23	1qbt	-14.49	-14.41	-14.61	-14.78	9896
24	2upj	-10.14	-10.20	-10.70	-9.49	7864
25	9hvp	-11.38	-11.43	-13.90	-13.49	10292
Average :		**-11.70**	**-11.15**	**-12.05**	**-11.68**	**8741**

5.5. Correlation between experimentally determined and calculated Binding energy

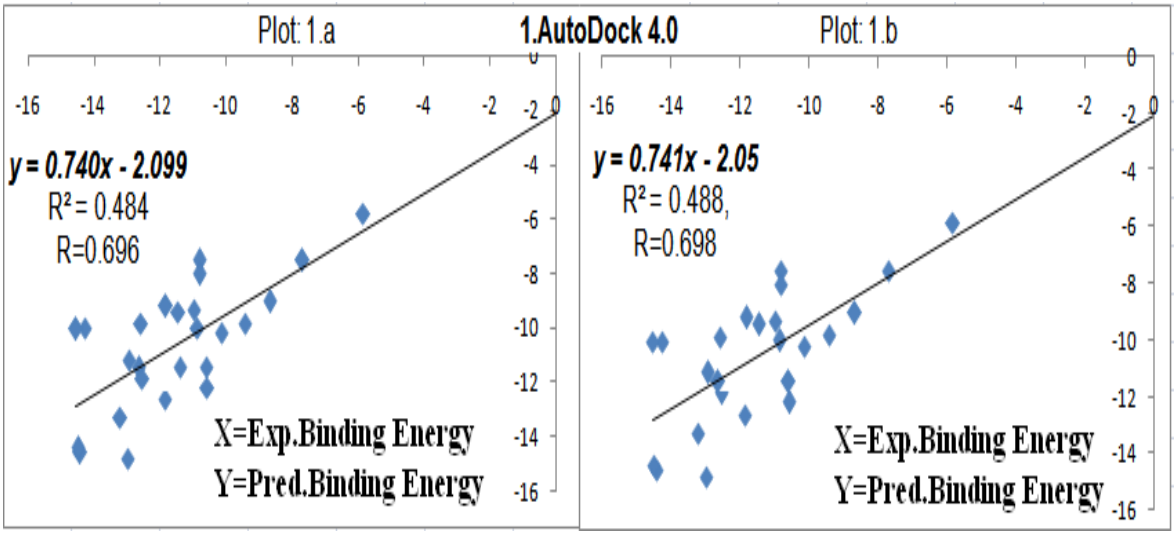
The primary objective of this study was to determine the binding energies by various docking methods and compared with experimental binding energies (table.6). Moreover, Table.7 shows the correlation coefficient of the experimentally determined and calculated binding energies from AutoDock 4.0 for the twenty-five protease-inhibitor complexes after protein rigid ligand flexible docking was 0.484 and the correlation coefficient from pardock was 0.801.

Table.7: correlation between experimental to predicted binding energies in autodock and pardock.

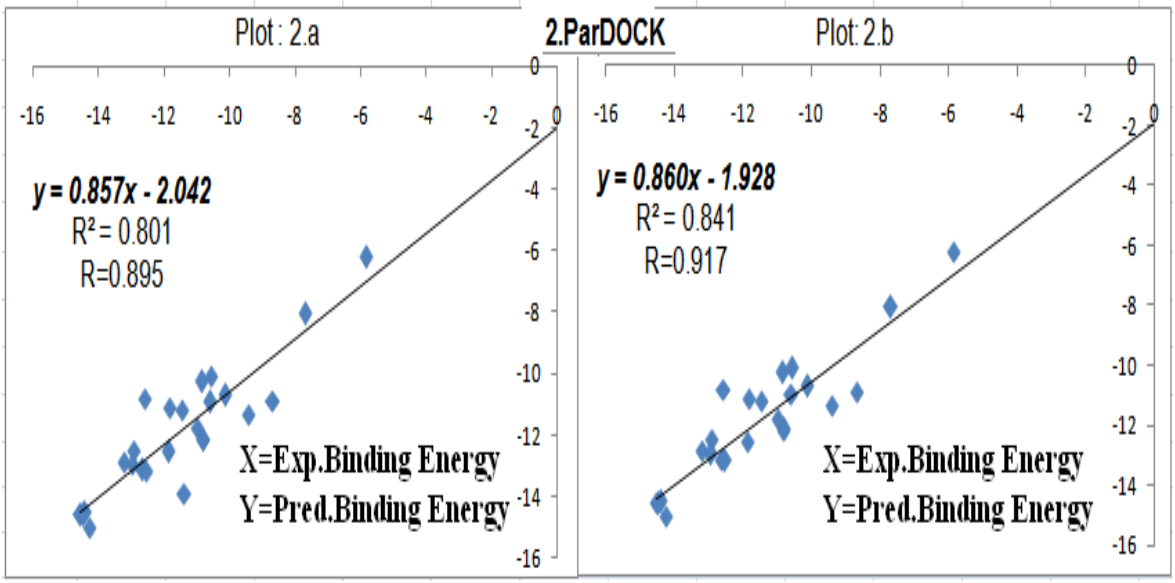
SNO	**Docking method**	**Correlation coefficient**		**Correlation**	
		initial	After removed one outlier	initial	After removed one outlier
1	Autodock4.0	0.484	0.488	0.696	0.698
2	Pardock	0.801	0.841	0.895	0.917

Experimentally determined vs. calculated binding energies for the twenty-five HIV-1 protease-inhibitor complexes

Plot 1: Graphs were plotted between experimental and predicted binding energies by Autodock 4.0 program. Correlation coefficient is 0.484(plot.1.a), changed to 0.488(plot.1.b) after outlier (9hvp) is eliminated in autodock4.0.



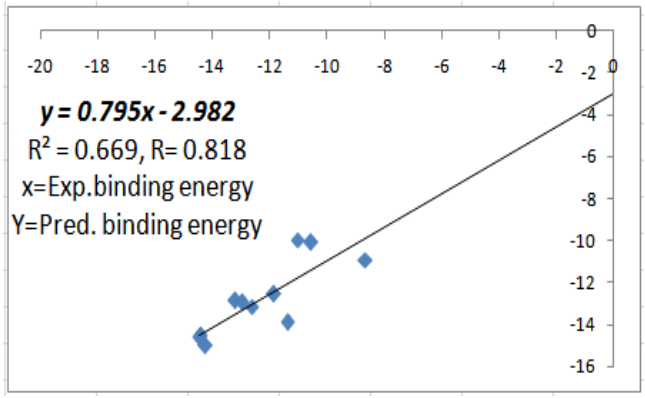
Plot 2: Graphs were plotted between experimental and predicted binding energies by Pardock program. Correlation coefficient is 0.801(plot.2.a), changed to 0.841(plot.2.b) after outlier (9hvp) is eliminated in pardock.



The best correlation coefficient 0.841 was observed in the pardock method after one outlier eliminated from the plots. The 1hiv predicted binding energy values are -11.47 kcal/mol in autodock,-13.14 kcal/mol in pardock. The 1hiv average predicted binding energy values in autodock and pardock was -12.30 kcal/mol, which is approximately equal to 1hiv experimental binding energy value -12.64 kcal/mol (table.6). The average binding energy value of these two predicted methods is approximately equal to its experimental binding energy value. The 1hiv shows consensus rank in autodock and pardock methods.

Patch dock score:

- Patchdock score ≥ 8741 having good binding solutions were observed(R≈0.82)
- 11 HIV-1PR-Inhibitor complexes having patchdock score ≥ 8741 out of 25
- Having Correlation (R)is ≈0.82for 11 HIV-1PR-Inhibitor complexes in pardock



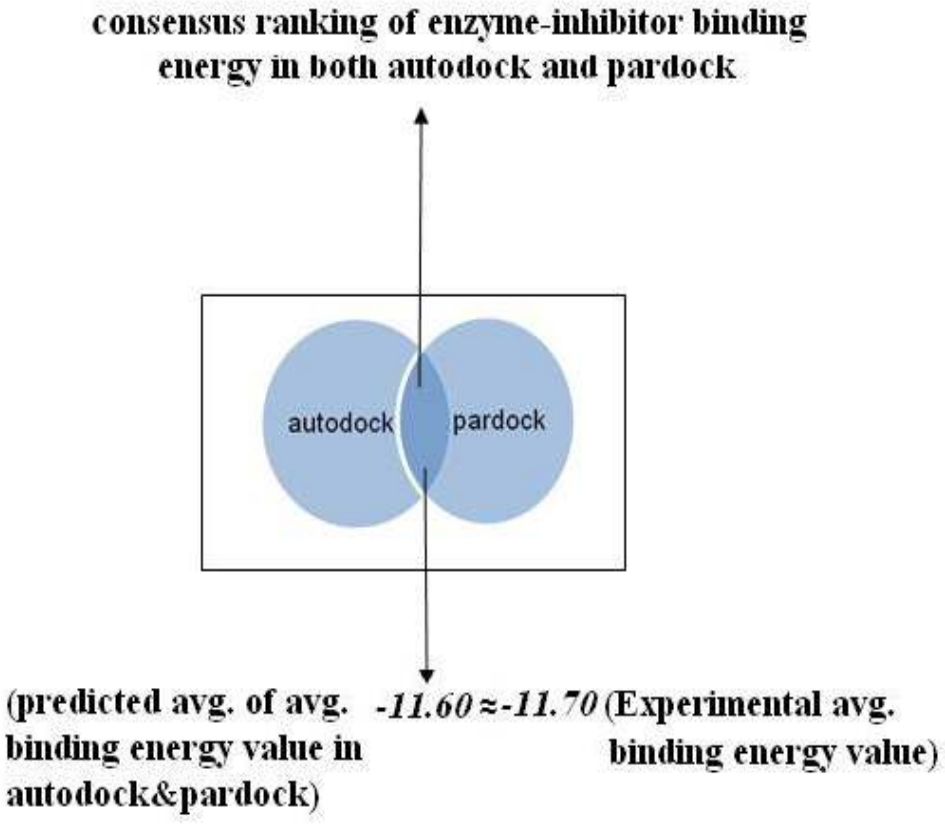
Plot.3: Patchdock score ≥ 8741 for 11 complexes.

Comparison of Average binding energy value:

In this study, the average experimental binding energy value for twenty-five HIV-1 Protease-inhibitor complexes shows -11.70 kcal/mol. which is approximately equal to the average of average predicted binding energy values of autodock and pardock was -11.60 kcal/mol (Fig.13). it is clearly indicates that consensus ranking of enzyme-inhibitor binding energy values in autodock and pardock methods are highly correlated to experimental binding energy values.

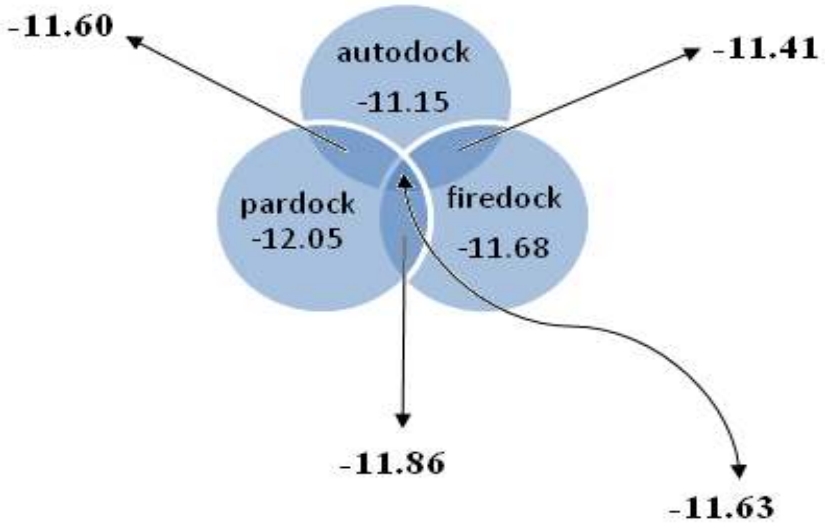
Here -11.70 is average experimental binding energy value and -11.60 are average of average predicted binding energy values in autodock and pardock method.

Fig.13: Ven diagram: consensus ranking of enzyme-inhibitor binding energy in both autodock and pardock.(In our study 1hiv complex consensus ranked on the trend lines in the plots).



In this study, the average experimental binding energy value for twenty-five HIV-1 Protease-inhibitor complexes shows -11.70 kcal/mol. Which is approximately equal to the average of average, predicted binding energy values of autodock, pardock and firedock was -11.63 kcal/mol (Fig.14). it is clearly indicates that consensus ranking of enzyme-inhibitor binding energy values in autodock, pardock and firedock methods are highly correlated to experimental binding energy values.

Fig.14: Ven diagram: consensus ranking of enzyme-inhibitor binding energy in autodock, pardock and firedock. (In our study 1hiv complex consensus ranked on the trend lines in the plots).



Here -11.70 is average experimental binding energy value and -11.60 are average of average predicted binding energy values in autodock and pardock. -11.86 and -11.41 are average of average predicted binding energy in pardock, firedock and autodock, firedock methods.

-11.63 is average of average predicted binding energy values in autodock, pardock and firedock,which is approximately equal to -11.70(average experimental binding energy values).

When a constant value 2.0 added to the predicted energies after docking with pardock, the binding energies of almost of the predictions were within -10 kcal/mol of the experimental values. Among these, 1hvr, 1hiv, 2upj are matched on the trend line in the plots between experimental and predicted binding energies in both autodock and pardock programs. firedock predicted binding energy values were very near to experimental binding energy values if 5.55 factor is multiple for the predicted binding energy values

Chapter 6

CONCLUSIONS

&

FUTURE WORK

6. CONCLUSIONS

In this study, we illustrate the importance of considering various docking methods to predict the structure and energetic of protein-ligand interactions. It is clear that pardock method for molecular docking simulations is produced good correlations between calculated and experimental binding energies. The average binding energy values of these two predicted methods was approximately equal to its average experimental binding energy value. Moreover, the average binding energy values of autodock, pardock and firedock predicted methods are approximately equal to its experimental binding energy value. The 1hiv shows consensus rank in autodock, pardock and firedock methods. The consensus ranking of enzyme-inhibitor complexes in various docking methods improve the binding energy predictions. One among twenty-five enzyme-inhibitor complexes, 1hiv shows consensus rank in autodock, pardock and firedock methods. *Thr*ee (1hvr, 1hiv and 2upj) enzyme-inhibitor complexes were observed in the plots of both autodock and pardock with more precision. The differences in these correlations may reflect various energy functions in different docking methods and biological features of the interactions of HIV-1 protease-inhibitor complexes. This approach will be helpful in computer assisted drug design.

FUTER WORK

Future work with larger data sets, various energy functions, various docking and binding energy evaluation methods, and more starting seeds, is necessary to determine the optimal parameters to robustly predict protein substrate binding energies in silico.

Chapter 7

REFERENCES

7. REFERENCES

1. Ashraf Brik and Chi-Huey Wong. HIV-1 protease: mechanism and drug discovery *Org.Biomol. Chem. (2003)* **1**, 5 – 14.

2. N. E. Kohl, E. A. Emini, W. A. Schleif, L. J. Davis, Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 4686.

3. (*a*) R. A. Kramer, M. D. Schaber, A. M. Skalka, K. Ganguly, F. Wong-Staal and E. P. Reddy, TLV-III gag protein is processed in yeast cells by the virus pol-protease. *Science*, 1986, 231, 1580; (*b*) J. R. Huff, HIV protease: a novel chemotherapeutic target for AIDS. *J. Med. Chem.*, 1991, **34**, 2305.

4. M. A. Navia, M. D. P. Fitzgerald, B. M. Mckeever, C.-T. Leu, J. C. Heimbach, W. K. Herber, I. S. Sigal, P. L. Darke and J. P. Spronger, *Thr*ee-dimensional structure of *Asp*artyl protease from human immunodeficiency virus HIV-1. *Nature*, 1989, **615**, 337.

5. A. Wlodawer, M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M.Selk, L. Clawson, J. Schneider and S. B. H. Kent, Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, 1989, **245**, 616.

6. T. Lee, G. S. Laco, B. E. Torbett, H. S. Fox, D. L. Lerner, J. H. Elder and C.-H. Wong, ProSelk, L. Clawson, J. Schneider and S. B. H. Kent, Analysis of the S3 and S3' subsite specificities of feline immunodeficiency virus (FIV) protease: development of a broad-based protease inhibitor efficacious against FIV, SIV, and HIV in vitro and ex vivo *Science*, 1989, **245**, 616.

7. S. C. Pettit, S. F. Michael and R. Swanstrom, The specificity of the HIV-1 protease. *Perspect. Drug Discovery Des.*, 1993, **1**, 69.

8. (*a*) J. Cairns, J. Overbought and S. Miller, 1988, **142**, 335; (*b*) F.W. Stahl, A unicorn in the garden*Nature Nature*, 1988, **142**, 112.

9. P. L. Darke, C. T. Leu, L. J. Davis, J. C. Heimbach, R. E. Diehl, W. S. Hill, R. A. Dixon and I.S. Sigal, Human immunodeficiency virus protease. Bacterial expression and characterization of the purified *Asp*artic protease *J. Biol. Chem.*, 1989, **264**, 2307.

10. S. Seelmeir, H. Schmidt, V. Turk and K. Von Der Helm, Human immunodeficiency virus has an *Asp*artic-type protease that can be inhibited by pepstatin. *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 6612.

11. J. Mous, E. P. Heimer and S. J. J. Le Grice, Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in Escherichia coli *J. Virol.*, 1988, **62**, 1433.

12. J. Hansen, S. Billich, T. Schulze, S. Sukrow and K. Moelling, Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. *EMBO J.*, 1998, **7**, 1785.

13. R. Lappatto, T. Blundell, A. Hemmings, J. Overington, A. Wilderspin, S. Wood, J. R. Merson, P. J. Whittle D. E. Danley, K. F. Geoghegan, S. J. Hawrylik, S. E. Lee, K. G. Scheld and P.Hobart, X-ray analysis of HIV-1 proteinase at 2.7 A^0 resolutions confirms structural homology among retroviral enzymes. *Nature*, 1989, **342**, 299.

14. (*a*) T. Hofmann, B. M. Dunn and A. L. Fink, A new chromophoric substrate for penicillopepsin and other fungal *Asp*artic pretenses. *Biochemistry*, 1984, **23**, 6956; (*b*) J. S. Fruton, Hydrolysis and Transpeptidation of Peptide Substrates by Acetyl-Pepsin*Adv. Enzyme.*, 1976, **44**, 1.

15. M. N. G. James and A. R. Sielecki, Stereo chemical analysis of peptide bond hydrolysis catalyzed by the *Asp*artic proteinase penicillopepsin. *Biochemistry*, 1985, **24**, 1.

16. K. Suguna, E. A. Padlan, C. W. Smith, W. D. Carlson and D. R. Davies, Binding of a reduced peptide inhibitor to the *Asp*artic proteinase from Rhizopus chinensis: implications for a mechanism of action. *Proc. Natl. Acad Sci. USA*, 1987, **84**, 6612

17.T. Hofmann, R. S. Hodges and M. N. G. James, Effect of pH on the activities of penicillopepsin and Rhizopus pepsin and a proposal for the productive substrate binding mode in penicillopepsin *Biochemistry*, 1984, **23**, 635

18. P. G. Schmidt, M. W. Holladay, F. G. Salituro and D. H. Rich, Identification of oxygen nucleophiles in tetrahedral intermediates: deuterium and oxygen-18 induced isotope shifts in carbon-13 NMR spectra of pepsin-bound peptide ketone pseudo substrates *Biochem. Biophys. Res. Commun.*, 1985, **129**, 597.

19. M. Jaskólski, A. G. Tomasselli, T. K. Sawyer, D. G. Staples, R. L. Heinrikson, J. Schneider,S. B. H. Kent and A. Wlodawer, Structure at 2.5-Å resolution of chemically

synthesized Human Immunodeficiency Virus Type 1 protease complexed with a hydroxyethylene-based inhibitor *Biochemistry*, 1991, **30**, 1600.

20. (*a*) M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshal, L. Clawsen, L.Selk, S. B. H. Kent and A. Wlodawer, X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Science*, 1989, 246, 1149;

(*b*) A. L. Swain, M. Miller, J. Green, D. H. Rich, J. Schneider, S. B. H. Kent and A. Wlodawer, X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor.*Proc. Natl. Acad. Sci.USA*, 1990, **87**, 8805.

21. L. J. Hyland, T. A. Tomaszek, Jr., G. D. Roberts, S. A. Carr, V. W. Magaard, H. L. Bryan, S. A. Fakhoury, M. L. Moore, M. D. Minnich, J. S. Culp, R. L. Desjarlais and T. D. Meek,Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by oxygen-18 isotope exchange. *Biochemistry*, 1991, **30**, 8441.

22. D. H. Slee, K. L. Laslo, J. H. Elder, I. R. Ollmann, A. Gustchina, K. Kervinen, A. Zdanov, A.Wlodawer and C.-H. Wong, Selectivity in the Inhibition of HIV and FIV Protease: Inhibitory and Mechanistic Studies of Pyrrolidine-Containing a-Keto Amide and Hydroxyethylamine Core Structures *J. Am. Chem. Soc.*, 1995, **117**, 11867.

23. W. E. Harte and D. L. Beveridge, Jr., Prediction of the protonation state of the active site *Asp*artyl residues in HIV-I protease-inhibitor complexes via molecular dynamics simulation. *J. Am. Chem. Soc.*, 1993, **115**, 3883.

24. L. H. Pearl and T. L. Blundell, The active site of *Asp*artic pretenses. *FEBS Lett.*, 1984, **174**, 96.

25. L. J. Hyland, T. A. Tomaszek and T. D. Meek, Human immunodeficiency virus-1 protease. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, 1991, **30**, 8454.

26. R. Smith, I. M. Brereton, R. Y. Chai and S. B. H. Kent, Ionization states of the catalytic residues in HIV-1 protease *Nat. Struct. Biol.*, 1996, **3**, 946.

27. Y. X. Wang, D. I. Freedberg, T. Yamazaki, P. T. Wingfield, S. J. Stahl, J. D. Kaufman, Y.Kiso and D. A. Torchia, Solution NMR Evidence That the HIV-1 Protease Catalytic *Asp*artyl Groups Have Different Ionization States in the Complex Formed with the Asymmetric Drug KNI-272. *Biochemistry*, 1996, **35**, 9945.

28. T. Yamazaki, L. K. Nicholson, D. A. Torchia, P. T. Wingfield, S. J. Stahl, J. D. Kaufman, C. J. Eyermann, C. N. Hodge, P. Y. S. Lam, Y. Ru, P. K. Jadhav, C. H. Chang and P. C. Weber, NMR and X-ray Evidence That the HIV Protease Catalytic *Asp*artyl Groups Are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-Based Inhibitor. *J. Am. Chem. Soc.*, 1994, **116**, 10791.

29. S. Piana, D. Sebastiani, P. Carloni and M. Parrinello, Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site. *J. Am. Chem. Soc.*, 2001, **123**, 8730.

30. T. D. Meek, B. D. Dayton, B. W. Metcalf, G. B. Dreyer, J. E. Strickler, J. G. Gorniak, M. Rosenberg, M. L. Moore, V. W. Magaard and C. Debouck, Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric *Asp*artic protease. *Proc. Natl. Acad. Sci. USA*, 1989, 86, 1841.

31. C. S. Lee, N. Choy, C. Park, H. Choi, Y. C. Son, S. Kim, J. H. Ok, H. Yoon and S. C. Kim, Design, synthesis, and characterization of dipeptide isostere containing cis-epoxide for the irreversible inactivation of HIV protease. *Bioorg. Med. Chem. Lett.*, 1996, **6**, 589.

32. T. D. Meek, E. J. Rodriguez and T. S. Angeles, Use of steady state kinetic methods to elucidate the kinetic and chemical mechanisms of retroviral protease. *Methods Enzyme.* 1994, 241, 127.

33. R. Bott, E. Subramanian and D. R. Davies, *Thr*ee-dimensional structure of the complex of the Rhizopus chinensis carboxyl proteinase and pepstatin at 2.5 Å resolution *Biochemistry*, 1982, **21**, 6956.

34. Renxiao Wang, Liang Liu, Luhua Lai, and Youqi Tang. SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. *Mol. Model.* 1998, *4.*

35. Samudrala R, Jenwitheesuk E (2003). Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations. *BMC Str. Biol.* **3**:2.

36. Gupta, A. Gandhimathi, A. Sharma, P. and Jayaram, B. Paddocks: All Atom Energy Based Monte Carlo Docking Protocol for Protein-Ligand Complexes. *Protein and Peptide Letters.2007.* **14**(7), 632-646.

37. Luty B.A, Wasserman Z.R, Stouten P.F.W (1995) Molecular Mechanics/Grid Method for the Evaluation of Ligand-Receptor Interactions. *J. Comp.Chem.* ***16*****:** 454-464.

38. Holloway M.K, Wei J.M (1995) A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J. Med. Chem.* ***38*****:** 305-317.

39. Ilya A.Vakser, omar G. *et al.* (1999) A systematic study of low-resolution recognition in Protein–protein complexes. *Proc natl. Acad. Sci. USA* **96**: 8477-8482.

40. Miklos Feher, Consensus scoring for protein-ligand interaction. D*rug discovery today*, **11:**9-10. *Nat. Struct. Biol.* 2006.**4:**8.

41. Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998), Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function *J. Computational Chemistry*, **19**: 1639-1662.

42. Huey, R., Morris, G. M., Olson, A. J. and Goodsell, D. S. (2007), A Semi empirical Free Energy Force Field with Charge-Based Desolvation. *J. Computational Chemistry*, **28**: 1145-1152.

43. DuhovnyD, Nussinov R. and Wolfson H.J (2002). Efficient unbound docking of rigid molecules. In GuigoR and Gusfield D. (eds), Proceedings of the Fourth International Workshop on Algorithms in Bioinformatics. Springer-Verlag GmbH Rome, Italy, September 17–21, 2002, Vol. 2452, pp. 185–200.

44. Connolly M.L. (1983) Solvent-accessible surfaces of proteins and nucleic acids. Science, 221, 709–713.

45. ConnollyM.L. (1983). Analytical molecular surface calculation. J. Appl. Crystallogr., 16, 548–558.

46. ZhangC, Vasmatzis, G., Cornette, J.L. and DeLisi, C. (1997) Determination of atomic desolvation energies from the structures of crystallized proteins. J. Mol. Biol., 267, 707–726.

47. N. Andrusier, R. Nussinov and H. J. Wolfson. FireDock: Fast Interaction Refinement in Molecular Docking. Proteins 2007, 69(1):139-59.

48. D. Duhovny, R. Nussinov, and H. J. Wolfson. Efficient unbound docking of rigid molecules. In R. Guigo and D. Gusfield, editors, Workshop on Algorithms in Bioinformatics, volume 2452, pages 185-200. Springer Verlag, 2002.

49. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson. PatchDock and SymmDock: zAcids. Res. 33: W363-367, 2005.

50. C. L. Kingsford, B. Chazelle, and M. Singh. Solving and analyzing side-chain positioning problems using linear and integer programming. Bioinformatics, 21(7):1028-1036, 2005.

51. C. Zhang, G. Vasmatzis, J. L. Cornette, and C. DeLisi. Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol, 267(3):707-726, April 1997.